UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/553,956	04/21/2000	Thomas A. Runkler	50277-452	7423

7590 06/13/2005

Stephen C Carlson
Ditthavong & Carlson P C
10507 Braddock Road
Suite A
Fairfax, VA 22032

EXAMINER	
PHAM, HUNG Q	
ART UNIT	**PAPER NUMBER**
2162	

DATE MAILED: 06/13/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

MAILED

JUN 13 2005

Technology Center 2100

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 09/553,956
Filing Date: April 21, 2000
Appellant(s): RUNKLER ET AL.

Margo Livesay, Reg. 41,946 & Stephen C. Carlson, Reg. 39,929
For Appellant

EXAMINER'S ANSWER

This is in response to the Appeal Brief filed 01/21/2005 and Supplemental Appeal Brief filed 03/30/2005.

S.O.J.

(1) Real Party in Interest

A statement identifying the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The brief does not contain a statement identifying the related appeals and interferences which will directly affect or be directly affected by or have a bearing on the decision in the pending appeal is contained in the brief. Therefore, it is presumed that there are none. The Board, however, may exercise its discretion to require an explicit statement as to the existence of any related appeals and interferences.

(3) Status of Claims

The statement of the status of the claims contained in the brief is correct.

(4) Status of Amendments After Final

The amendment after final rejection filed on 03/30/2005 has been entered.

(5) Summary of Invention

The summary of invention contained in the brief is correct.

(6) Issues

The appellant's statement of the issues in the brief is correct.

(7) Grouping of Claims

The rejection of claims 1-6, 10, 12-14, 16-23, 27, 29-31 and 33-36 stand or fall together because appellant's brief does not include a statement that this grouping of claims does not stand or fall together and reasons in support thereof. See 37 CFR 1.192(c)(7).

(8) Claims Appealed

The copy of the appealed claims contained in the Appendix to the brief is correct.

(9) Prior Art of Record

- 6,247,016 B1 Rastogi et al. 06-2001
- Shimoji et al., "Data Clustering with Entropical Scheduling", 1994 IEEE International Conference, 27 June-2 July 1994, vol. 4, pages 2423-2428.
- Hall et al., "Generating Fuzzy Rules from Data", Proceedings of the Fifth IEEE International Conference, 08-11 Sept. 1996, vol. 3, pages 1757-1762.
- Shafer et al., "SPRINT: A Scalable Parallel Classifier for Data Mining", Proceedings of the 22nd VLDB Conference Mumbai (Bombay), India, 1996, pages 544-555.
- Janikow, C.Z., "Fuzzy Decision Trees: Issues and Methods", Systems, Man and Cybernetics, Part B, IEEE Transactions, Feb. 1998, Vol. 28, pages 1-14.

- Chow et al., "On The Optimal Choice of Parameters in a Fuzzy C-Means Algorithm", Fuzzy Systems, 1992., IEEE International Conference, 8-12 March 1992, pages 349-354.

(10) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

Claims 17 and 34-36 are rejected under 35 U.S.C. 102(a) as being anticipated by Applicant Admitted Prior Art [Background Of The Invention, pages 1-5].

Regarding claims 17 and 34, in the background, FID3 is a conventional ***computer implemented method for generating a decision tree for a plurality of data characterized by a plurality of features*** (page 3, line 25-page 4, line 22). FID3 technique comprising:

performing a plurality of fuzzy cluster analysis along each of the features to calculate a maximal partition coefficient and a corresponding set of one or more fuzzy clusters, said

maximal partition coefficient corresponding to one of the features (as illustrated from page 3, line 25-page 4, line 22, membership function between 0.0 and 1.0 as ***a fuzzy cluster analysis*** is used to represent the degree to which the object belongs to the class; patients' features, e.g., age, temperature, are grouped into Young, Old, Normal, Feverish as ***fuzzy cluster***, using membership function, e.g., $\mu_{young}(2) = 0.99$, $\mu_{old}(2) = 0.01$, $\mu_{young}(65) = 0.13$, $\mu_{old}(2) = 0.87$, and a test $\mu_{young}(X_i) < 0.5$ to maximize information gain or ***maximal partition coefficient***, as further disclosed in the Background at page 3, lines 15-17, information gain for discriminating objects at branch node or ***partition coefficient*** is calculated by finding average entropy of each feature);

selecting the one of the features corresponding to the maximal partition coefficient (Background, page 3, lines 15-17);

building the decision tree based on the corresponding set of one or more fuzzy clusters (Background, page 4, lines 15-22).

Regarding claims 35 and 36, the admission teaches all the claimed subject matters as discussed in claims 17 and 34, the admission further discloses ***the maximal partition is based on membership functions of the data for the set of one or more clusters*** (page 4, lines 10-15).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Claims 1-3 and 18-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Rastogi et al. [USP 6,247,016 B1] in view of Shimoji et al. [Data Clustering with Entropical Scheduling].

Regarding claims 1 and 18, Rastogi teaches a method and a computer readable medium bearing instruction for classifying data using a decision tree. As shown in FIG. 1, there is a single record corresponding to each loan request, characterized two attributes, salary and education level completed (Col. 2, lines 50-56).

As shown in FIG. 2, `salary` is selected ***from among the features characterizing the data associated with*** the root node, and the test `is the salary level of the applicant less than $20,000.00` (Col. 2, lines 62-63) is to split the root node N

into N_1 and N_2 (FIG. 3, line 8). The test is based on the process of calculation of the least entropy by scanning the attribute list from the beginning to calculate an entropy for each split point or each numeric attribute in order to determine the least entropy (Col. 4, lines 25-52). In short, the technique as discussed indicates the steps of ***selecting a feature from among the features characterizing the data associated with the node***, and the process of determining the least entropy as ***performing a cluster analysis along the selected feature to group the data into one or more cluster***.

The left arc that connects the root node to node 30 is labeled *YES* indicating that node 30 is to be reached if the *salary < $20,000*. On the other hand, the right arc connects root node to another branch node is labeled *NO* indicating the branch node is to be reached if *salary > $20,000*. The branch node is labeled *ACCEPT* (FIG. 2). This performs the claimed ***constructing one or more arcs of the decision tree at the node respectively for each of the one or more clusters***.

As in FIG. 1, the first applicant has a salary of $15,000. Thus, at root node 10, the condition yields a *YES*, the attributes of this first applicant are passed on to the left branch, where an additional test takes place. If the condition resulted in a *NO* answer, the attribute of this applicant would have been passed to the right branch and leaf 20 would have been formed, classifying this applicant in the class of applicants whose loan request is accepted (Col. 3, lines 46-58). As seen, the attributes of first record are passed to the left branch to node 30 characterized by *Education* feature for another test, and the attributes of second record are passed to the right branch to node 20 characterized by *ACCEPT* attribute as the step of ***projecting the data in each of the clusters,***

wherein the projected data are characterized by the plurality of the features but for the selected feature. As shown in FIG. 3 is the procedure to build the decision tree. A loop is set up at line 3, the root node is queued at line 2 and de-queued at line 4, root node is split into nodes 30 and 20 at line 8, appended to the queue at line 9 (FIG. 3). The procedure is recursively performed on node 30 at line 3 with another process of calculation of the least entropy and another test for `Education` as the selected and projected feature (Col. 3, lines 5-9). As seen, the procedure of building decision tree with a loop as discussed indicates the step of ***recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the cluster.***

Rastogy does not explicitly teach the cluster analysis is ***based on distances between the data and respective one of more centers of the one or more clusters.***

Shimoji discloses a method of clustering a set of data by using a clustering error ***based on distances between the data and respective one of more centers of the one or more clusters*** (Shimoji, Introduction). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to combine clustering error as taught by Shimoji to analyze a cluster when grouping data into one or more cluster of a decision tree.

Regarding claims 2 and 19, Rastogi and Shimoji teaches all the claimed subject matters as discussed in claims 1 and 18, Rastogi further discloses the steps of ***performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure, said maximal cluster validity measure corresponding to one of the***

features; and selecting the one of the features that corresponds to the maximal cluster validity measure (Col. 4, lines 25-52).

Regarding claims 3 and 20, Rastogi and Shimoji teaches all the claimed subject matters as discussed in claims 2 and 19, Rastogi further discloses the step: *for each of the features, performing a plurality of cluster analyses along said each of the features for a plurality of cluster numbers to calculate respective partition coefficients; and determining the maximal cluster validity measure from among the partition coefficients* (Col. 4, lines 25-52).

Claims 4 and 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Rastogi et al. [USP 6,247,016 B1] in view of Shimoji et al. [Data Clustering with Entropical Scheduling] and Applicant Admitted Prior Art [Background Of The Invention, pages 1-5].

Regarding claims 4 and 21, Rastogi teaches all the claimed subject matters as discussed in claims 1 and 18, but fails to disclose the step of *performing the cluster analysis includes the step of performing a fuzzy cluster analysis*. Applicant Admitted Prior Art teaches the technique of using fuzzy cluster analysis for a decision tree (page 4, lines 1-5). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Rastogi method by using fuzzy cluster analysis for a decision tree as taught in the admission in order to calculate the maximizing information gains.

Claims 5 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Rastogi et al. [USP 6,247,016 B1] in view of Shimoji et al. [Data Clustering with Entropical Scheduling], Applicant Admitted Prior Art [Background Of The Invention, pages 1-5] and Hall et al. [Generating Fuzzy Rules from Data].

Regarding claims 5 and 22, Rastogi and Applicant Admitted Prior Art teaches all the claimed subject matters as discussed in claims 4 and 21, but fails discloses the step of *performing the fuzzy cluster analysis includes the step of performing a fuzzy c-means analysis*. Hall teaches the technique of using fuzzy c-means for a decision tree (Hall, Generating Fuzzy Rules from Data). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Rastogi and Applicant Admitted Prior Art method by including the technique of using fuzzy c-means in order to determine the number of cluster.

Claims 6 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Rastogi et al. [USP 6,247,016 B1] in view of Shimoji et al. [Data Clustering with Entropical Scheduling] and Shafer et al. [SPRINT: A Scalable Parallel Classifier for Data Mining].

Regarding claims 6 and 23, Rastogi teaches all the claimed subject matters as discussed in claims 1 and 18, but fails to disclose the step of *performing the cluster*

analysis includes the step of performing a hard cluster analysis. Shafer teaches a method of forming a decision tree by performing a hard cluster analysis (Shafer, SPRINT: A scalable Parallel Classifier for Data Mining, pages 544-550, especially Abstract and Introduction pages 544-545). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Rastogi method by including the technique of hard cluster analysis in order to optimize the system by using a regular cluster for classifying records of unknown class.

Claims 1-5 and 18-22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Janikow [Fuzzy Decision Trees: Issues and Method] in view of Choe et al. [On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm].

Regarding claims 1 and 18, Janikow teaches method of building a fuzzy decision tree. To simplify the method, FIG. 4 & 5 on pages 8-9 could be used to illustrate the Janikow method.

As illustrated at Procedure to Build a Fuzzy Decision Tree (step 4, page 7:2) and FIGS. 4, 8, `Employment` at the root node as ***a selected feature from among the features***, e.g., `Employment, Income`, ***characterizing the data associated with the node*** for ***constructing*** Low, Medium and High, which are ***one or more arcs of the decision tree at the node respectively for each of the one or more clusters***, and

projecting the data in each of the clusters, wherein the projected data are characterized by the plurality of the features but for the selected feature (Janikow, page 8:2, the root R gets expanded with three children as shown in FIG. 4).

Janikow further discloses the step of ***performing a cluster analysis along the selected feature to group the data into one or more clusters*** (Procedure to Build a Fuzzy Decision Tree, step 4, page 7:2), and

recursively performing the steps of selecting a feature and performing the cluster analysis on the projected data in each of the clusters (Procedure to Build a Fuzzy Decision Tree, step 4, page 7:2, as suggested by Janikow, step 4 is performed at each node of the expanded tree,).

Janikow does not explicitly illustrate the cluster analysis is ***based on distances between the data and respective one or more centers of the one or more cluster***.

Choe discloses a Fuzzy C-Means Algorithm to maximize the number of data points in a cluster by using a fuzzy constraint. The Choe cluster analysis is ***based on distances between the data and respective one or more centers of the one or more cluster*** (Choe, Fuzzy C-Means Algorithm, pages 350-351).

It would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Janikow method by using the error constraint based on the distance between data and center of cluster to build a decision tree in order to maximize the number of data points in a cluster.

Regarding claims 2 and 19, Janikow and Choe teaches all the claimed subject matters as discussed in claims 1 and 18, Choe further discloses the steps of ***performing a plurality of cluster analyses along each of the features to calculate a maximal cluster validity measure, said maximal cluster validity measure corresponding to one of the features; and selecting the one of the features that corresponds to the maximal cluster validity measure*** (Choe, Fuzzy C-Means Algorithm, pages 350-351).

Regarding claims 3 and 20, Janikow and Choe teaches all the claimed subject matters as discussed in claims 2 and 19, Choe further discloses the step ***performing a plurality of cluster analyses along said each of the features for a plurality of cluster numbers to calculate respective partition coefficients; and determining the maximal cluster validity measure from among the partition coefficients*** (Choe, Fuzzy C-Means Algorithm, pages 350-351).

Regarding claims 4 and 21, Janikow and Choe teaches all the claimed subject matters as discussed in claims 1 and 18, Janikow further discloses the step of ***performing the cluster analysis includes the step of performing a fuzzy cluster analysis*** (Janikow, page 6, Fuzzy Decision Tree).

Regarding claims 5 and 22, Janikow and Choe teaches all the claimed subject matters as discussed in claims 4 and 21, Choe further discloses the step of ***performing***

the fuzzy cluster analysis includes the step of performing a fuzzy c-means analysis (Choe, Fuzzy C-Means Algorithm, pages 350-351).

Claims 6 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Janikow [Fuzzy Decision Trees: Issues and Method] in view of Choe et al. [On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm] and Shafer et al. [SPRINT: A Scalable Parallel Classifier for Data Mining].

Regarding claims 6 and 23, Janikow and Choe teaches all the claimed subject matters as discussed in claims 1 and 18, but fails to disclose the step of *performing the cluster analysis includes the step of performing a hard cluster analysis*. Shafer teaches a method of forming a decision tree by performing a hard cluster analysis (Shafer, SPRINT: A scalable Parallel Classifier for Data Mining, pages 544-550, especially Abstract and Introduction pages 544-545). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Janikow and Choe method by including the technique of hard cluster analysis in order to optimize the system by using a regular cluster for classifying records of unknown class.

Claims 10, 12, 16, 27, 29 and 33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Janikow [Fuzzy Decision Trees: Issues and Method].

Regarding claims 10 and 27, Janikow teaches a method for generating a decision tree for a plurality of data characterized by a plurality of features (TABLE 1, FIG. 4, page 8).

As illustrated by Janikow at step 4 of the procedure to build a Fuzzy Decision Tree on pages 7-8, a plurality of cluster analyses as in pages 7-8 along `Employment` and `Income` to calculate a plurality of information gain to split the node as ***partition coefficients*** (G^R_{Inc}, G^R_{Emp}), *Emp* is the selected attribute corresponds to G^R_{Emp} as ***a maximal partition coefficient*** from among the G^R_{Inc}, and G^R_{Emp} as partition coefficients. The root gets expanded with the following three children based on the selected Emp, and the decision tree is built based on the three children as in FIG. 4. In short, the Janikow technique of building the decision tree indicates the steps of ***performing a plurality of cluster analyses along each of the features to calculate a plurality of respective partition coefficients***, ***selecting the one of the features corresponding to a maximal partition coefficient from among the partition coefficients***; ***subdividing the data into one or more groups based on the selected feature***; and ***building the decision tree based on the one or more groups***.

Janikow does not explicitly teach the G^R_{Inc}, and G^R_{Emp} as ***the partition coefficients*** that are ***based on membership functions of the data for one or more clusters in respective said cluster analyses***. However, as disclosed by Janikow, at each node, the set of remaining attributes from $V - V^N$ is searched, I superscript S^N_{Vi} is calculated, and information gain as ***partition coefficient*** $G^N_i = I^N$ - I superscript S^N_{Vi}. As seen, obviously, G^N_i as ***partition coefficient*** depends on the value of I superscript S^N_{Vi} that is based on the function f_2 of

data for the corresponding cluster of FIG. 4 in respective cluster analyses. It would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Janikow method by using function f_2 as the membership function for calculating information gain as ***partition coefficient*** in order to split a node based on the attribute that has the highest information gain.

Regarding claims 12 and 29, Janikow teaches all the claimed subject matters as discussed in claims 10 and 27, Janikow further discloses the step of ***performing a plurality of fuzzy cluster analyses*** (pages 7-8).

Regarding claims 16 and 33, Janikow teaches all the claim subject matters as discussed in claims 10 and 27, Janikow further discloses the step of ***projecting the data in each of the group, wherein the projected data are characterized by the plurality of the features but for the selected feature; and recursively performing the steps of selecting a feature, comprising selecting a new one of the features corresponding to a new maximal partition coefficient and subdividing the data into one or more new groups based on the selected new feature*** (Janikow, pages 7-9, Procedure to Build a Fuzzy Decision Tree).

Claims 13 and 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Janikow [Fuzzy Decision Trees: Issues and Method] in view of Choe et al. [On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm].

Regarding claims 13 and 30, Janikow teaches all the claim subject matters as discussed in claims 10 and 27, Janikow does not explicitly teach the step of *performing the fuzzy cluster analyses includes the step of performing a plurality of fuzzy c-means analyses*. Choe discloses a Fuzzy C-Means Algorithm to maximize the number of data points in a cluster by using a fuzzy constraint (Choe, On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to use the fuzzy c-means analyses as taught by Choe in order to maximize the number of data points in a cluster.

Claims 14 and 31 are rejected under 35 U.S.C. 103(a) as being unpatentable over Janikow [Fuzzy Decision Trees: Issues and Method] in view of Shafer et al. [SPRINT: A Scalable Parallel Classifier for Data Mining].

Regarding claims 14 and 31, Janikow teaches all the claimed subject matters as discussed in claims 1 and 18, but fails to disclose the step of *performing the cluster analysis includes the step of performing a hard cluster analysis*. Shafer teaches a method of forming a decision tree by performing a hard cluster analysis (Shafer, SPRINT: A scalable Parallel Classifier for Data Mining, pages 544-550, especially Abstract and Introduction pages 544-545). Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to modify the Janikow method by

including the technique of hard cluster analysis in order to optimize the system by using a regular cluster for classifying records of unknown class.

Allowable Subject Matter

Claims 15 and 32 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Regarding to claims 15 and 32, the closet available prior arts, USP 6,247,016 B1, issued to Rastogi and Janikow (Fuzzy Decision Trees: Issues and Method) also teaches the technique of refining a node of a decision tree. However, as in claims 15 and 32, Rastogi and Janikow fails to teach or suggest the steps of ***calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data; determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster***. Therefore, the invention is allowable over the prior arts of record for being directed to a combination of claimed elements including the providing steps as indicated above.

Claims 7-8 and 24-25 are allowed.

Regarding to claims 7-8 and 24-25, the closet available prior arts, USP 6,247,016 B1, issued to Rastogi and Janikow (Fuzzy Decision Trees: Issues and Method) also

teaches the technique of refining a node of a decision tree. However, as in claims 7-8 and 24-25, Rastogi and Janikow fails to teach or suggest the steps of ***calculating a domain ratio of a difference in domains limits of the data over a difference in domain limits of a superset of the data; determining whether the domain ratio has a predetermined relationship with a predetermined threshold; and if the domain ratio has the predetermined relationship with the predetermined threshold, then grouping the data into a single cluster***. Therefore, the invention is allowable over the prior arts of record for being directed to a combination of claimed elements including the providing steps as indicated above.

(11) Response to Argument

A. <u>RESPONSE TO ARGUMENTS WITH RESPECT TO THE REJECTION OF CLAIMS 17 AND 34-36 UNDER 35 U.S.C § 102 AS BEING ANTICIPATED BY THE AMIDSSION.</u>

(1) As argued by appellants at page 7 with respect to the rejection of claims 17 and 35:

> *The Examiner's reasoning is predicated on the mistaken assumption that a maximal partition coefficient can be equated to a maximum information gain (Office Action of May 20, 2004, p. 16, emphasis original):*
>
> > *As seen, $\mu_{young}(Xi)$ and $\mu_{old}(Xi)$ as a plurality, of fuzzy cluster analyses is performed along each of the age features to calculate the highest information gain corresponding to one of the features as* ***maximums partition coefficient*** *and for two fuzzy sets Young and Old, then the attribute with the highest information gain is selected to discriminate objects at the branch node to build the decision tree based on two fuzzy sets Young and Old.*
>
> *However, the Background merely states at p. 4:13-14 that "As in 1D3, FID3 generates its decision trees by maximizing information gains." There is no support in the Background, admitted or otherwise, for the Examiner's glossing of "highest information gain" as a "maximum partition coefficient."*

Examiner respectfully traverses because of the following reason:

The Background does not only state at page 4, lines 13-14 a *highest information gain* but also discloses the claimed performing a plurality of cluster analyses along each of the features to calculate the highest information gain. As illustrated in the Background at page 3, line 25-page 4, line 5, and page 3, lines 15-16:

> *... fuzzy logic employs a "membership function" between 0.0 and 1.0 to represent the degree to which the object belongs to the class. Rather than categorize a patient's age as "twelve years and below" and "above twelve years," two fuzzy sets, Young and Old, can be employed, such that a two-year old may have a membership function in the Young fuzzy set $\mu_{young}(2) = 0.99$ but a membership function in the Old fuzzy set $\mu_{old}(2) = 0.01$* (Background, page 4, lines 1-5). *A branch node is created and the attribute with the highest information gain is selected if that attribute were used to discriminate objects at the branch node* (Background, page 3, lines 15-16). *In the example of FIG. 5, the arcs 512 and 514 emanating from branch node 510 could be fuzzified by a membership function on a Young fuzzy set and an Old fuzzy set, respectively. For example, arc 512 could be the test* $\mu_{young}(X_i)$ *< 0.5 or other values that maximizes the information gain* (Background, page 4, lines 15-19). *The information gain is calculated by finding the average entropy of each attribute* (Background, page 3, line 17).

As seen, membership function for representing the degree of grouping $\mu_{young}(X_i)$ and $\mu_{old}(X_i)$ as ***a plurality of fuzzy cluster analyses*** is ***performed along each of the features***, e.g., age, ***to calculate*** the highest information gain as ***maximal partition coefficient corresponding to one of the features*** via the test, e.g., $\mu_{young}(X_i) < 0.5$.

(2) As argued by appellants at page 7 with respect to the claimed ***maximal partition coefficient***:

> *The basis for the Examiner's highly unusual understanding appears to be a phrase in the Detailed Description of the Specification, p. 15:4, that explains a property of the partition coefficient as "which quantifies the goodness of the clustering" (Office Action, p. 3). This statement in the Detailed Description, however, is clearly not found in the Background or admitted prior art. Furthermore, quantifying the goodness of the clusters does not mean that any number that might have some connection to fuzzy clustering must be a partition coefficient.*

Examiner respectfully traverses.

As set forth in Manual of Patent Examining Procedure § 2111:

> *during patent examination, the pending claims must be given their broadest reasonable interpretation consistent with the specification.*

As defined in the specification, page 15, lines 3-4, *a partition coefficient, which quantifies the goodness of the clustering*. Thus, *a maximal partition coefficient* is considered as a number that indicates a highest measurement of division property. As disclosed in the Background, the ID3 is a recursive algorithm that starts with a set of training objects that belong to a set of predefined classes. If all the objects belong to a single class, then there is no decision to make and a leaf node is created and labeled with the class. Otherwise, a branch node is created and the attribute with the highest information gain is selected if that attribute were used to discriminate objects at the branch node. The information gain is calculated by finding the average entropy of each attribute (Background, page 3, lines 15-17). As seen, each attribute associates with a calculated information gain, and the attribute with the highest information gain is selected for branching the node if that attribute is able to discriminate objects. Thus, the highest information gain is equated with the *maximal partition coefficient* because it *quantifies the goodness of the clustering*.

(3) As argued by appellants at pages 7 and 8 with an excerpt of Janikow that teaches the technique of calculating the information gain to come to the conclusion:

> *one of ordinary skill in the art would not understand, based either on the Background or the prior art, that either ID3 or FID3 builds their decision trees using a maximal partition coefficient. In fact, such a person of ordinary skill would not even equate a maximal partition coefficient with a maximum information gain. Well-settled case law holds that the words of a claim must be read as they would be interpreted by those of ordinary skill in the art. In re Baker Hughes Inc., 215 F.3d 1297, 55 USPQ2d 1149 (Fed. Cir. 2000); In re Morris, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997). "Although the PTO must give claims their broadest reasonable interpretation, this interpretation must be consistent with the one that those skilled in the art would reach." In re Cortright, 165 F.3d 1353, 1369, 49 USPQ2d 1464, 1465 (Fed. Cir. 1999).*

Examiner respectfully traverses because a person of ordinary skill in the art would equate a maximal partition coefficient with a maximum gain. Specifically, the maximum gain ***corresponding to one of the features*** is ***calculated by performing a plurality of fuzzy cluster analysis along each of the features*** as discussed above, and the conventional maximum gain meets the requirement of the claimed ***maximal partition coefficient***.

B. RESPONSE TO ARGUMENTS WITH RESPECT TO THE REJECTION OF CLAIMS 1-6, 10, 12-14, 16, 18-23, 27, 29-31 and 33 UNDER 35 U.S.C § 103 AS BEING UNPATENABLE OVER *JANIKOW* AND OTHER APPLIED ART.

(1) As argued by appellants at page 10 with respect to claims 10, 12, 16, 27, 29 and 33:

> *The Office Action of May 21, 2004, p, 13, admits that "Janikow does not explicitly teach the G^R_{Ino} and G^R_{Emp} as the partition coefficients." Without any teaching of a partition coefficients, there is nothing in Janikow to teach or other suggest the following step of selecting and subdividing which recites the*

> *"partition coefficients," nor for that matter the next step of subdividing that is "based on the selected feature." In fact, the Examiner recognizes that Janikow uses another measure to split the node, viz., "to calculate a plurality of information gain to the split the node." As explained above in Section VII. A., one of ordinary skill in the art would not confuse information gain with a partition coefficient.*

Generally, as in the Office Action, examiner admitted Janikow does not explicitly teach the claimed membership functions, not the partition coefficient.

Specifically, the referred Section VII. A. from Janikow (page 5:2 to page 6:1) is summarized as below:

> *The root of the decision tree contains all training examples. It represents the whole description space since no restrictions are imposed. Each node is recursively split by partitioning its examples. A node becomes a leaf when either its samples come from a unique class or when all attributes are used on the path. When it is decided to further split the node, one of the remaining attributes (i.e., not appearing on the current path) is selected. Domain values of that attribute are used to generate conditions leading to child nodes. The examples present in the node being split are partitioned into child nodes according to their matches to those conditions. One of the most popular attribute selection mechanisms is one that maximizes information gain [25]. This mechanism, outlined below, is computationally simple as it assumes independence of attributes.*
>
> *1) Compute the information content at node N.*
>
> *2) For each attribute a_i not appearing on the path to N and for each of its domain values a_{ij}, compute the information content $I^{N|a_{ij}}$ in a child node restricted by the additional condition $a_i = a_{ij}$.*
>
> *3) Select the attribute a_i maximizing the information gain*
>
> $$I^N - \sum_j^{|D_i|} (w_j \cdot I^{N|a_{ij}})$$
>
> *4) Split the node using the selected attribute.*

The process of Section VII.A as referred by appellants is detailed as in page 7, wherein information gain for each of the attribute is more clearly defined as below:

$$G_i^N = I^N - I^{S_{V_i}^N}$$

and attribute v_i with a maximal information gain G_i^N is selected to build the decision tree.

As seen, information gains as ***partition coefficients for each of the*** attributes or ***features are calculated***, then the attribute or ***feature*** that has the maximum information gain as ***maximal partition coefficient is selected***, the root node containing all training examples as ***data is*** split into a plurality of sub-nodes or ***subdivided into one or more groups based on the selected*** attribute or ***feature***.

(2) As argued by appellants at page 10 with respect to claims 10, 12, 16, 27, 29 and 33:

> *Recognizing Janikow's deficiency, the Examiner contends that it would have been obvious "to modify the Janikow method by using function f2 as the membership function ... in order to split a node" (p. 13). However, Janikow, p. 9, expressly teaches against just such a modification: "To define the decision procedure, we must define f0, f1, f2, f3 for dealing with samples presented for classification. These operators may* ***differ from those used for tree building****-let us denote them g0, g1,g2, g3." Thus, Janikow discloses a distinction between classification functions, e.g. f2, and tree building functions, e.g. g2, and one of ordinary skill in the art would not be motivated to disregard Janikow's distinctions and principle of operation when making modifications of its method.*

Examiner respectfully traverses. Although, operators f_0, f_1, f_2, f_3 may differ from g_0, g_1, g_2, g_3 for tree building (Janikow, page 9:1), the purpose of these two operators is

to define the membership value of the membership function. As illustrated at pages 7:1, and 9:1, f_0 and g_0 operators have the same membership values as shown below:

$$f_0(e_j, [V_i \text{ is } v_p^i]) = \begin{cases} \frac{1}{|D_i|} & \text{if } u_j^i \text{ unknown} \\ \mu_{v_p^i}(u_j^i) & \text{otherwise.} \end{cases}$$

$$g_0(e_j, [V_i \text{ is } v_p^i]) = \mu_{v_p^i}(u_j^i).$$

As seen, the difference of the two operators is the ignoring unknown u^i_j of g_0. Other than that, two operators apply to the same values, and produce the same membership values. Therefore, the difference of operators does not affect the defined membership function $\mu_{v_p^i}(u_j^i)$.

(3) As argued by appellants at pages 11 and 12 with respect to claims 1-5 and 18-22:

> *Janikow does not show "recursively ... performing the cluster analysis." The Examiner's rejection, which merely cites pp. 7-9 without explanation, is inadequate, since Janikow discloses a distinction between classification functions, e.g. f2, and tree building functions, e.g. g2. In fact, by keeping classification and tree building distinct, Janikow teaches against "recursively ... performing the cluster analysis" in general and the proposed modification of Janikow to use Choe et al.'s classification system. Because of this distinction, Janikow actually teaches against using any classification function in Choe et al. for tree building (cf. claims 1 and 18: "constructing one or more arcs of the decision tree").*

Examiner respectfully traverses.

Firstly, the distinction of f_2 and g_2 does not affect the membership function as discussed above. Therefore, Janikow does not *teaches against using any classification function in Choe et al. for tree building* as asserted by appellants.

Secondly, Janikow teaches the root of the decision tree contains all training examples. Each node is recursively split by partitioning its examples (Janikow, page 6:1, first paragraph). In fuzzy set theory, a membership function $\mu_u(V) : U \rightarrow [0,1]$ represents the degree to which u € U belong to the set v. A fuzzy linguistic variable V is an attribute whose domain contains linguistic values or fuzzy terms, which are labels for the fuzzy subsets, e.g., *Low*, *Medium* and *High* as domain values for attribute *Income* (Janikow, page 2:2, second paragraph). The procedure to build a decision tree is described in section V, page 7, starting with all the examples E in the root node, and at any node N still to be expanded, P_k^N is computed for information gain G_i^N as below:

$$P_k^N = \sum_{j=1}^{|E|} f_2(\chi_j^N, \mu_{v_k}(y_j))$$

As seen, membership function $\mu_u(V)$ of an attribute u represents the degree to which u belong to v is a cluster analysis. Membership function is performed at any expanded node, and used for calculating P_k^N, and information gain G_i^N. In short, membership function as ***cluster analysis*** is ***recursively performed.***

As argued by appellants at page 12 with respect to claims 6 and 23:

> *Shafer et al., directed to a scalable parallel classifier for data mining (per title), discusses only "classes" of data and partitions of the data (e.g., p. 546, left column), and makes no mention of any "cluster analysis," much less "performing a hard cluster analysis."*

Examiner respectfully traverses. A hard clustering algorithm, as well known in the art, is to cluster data into non-overlapping groups. The Shafer technique, in order to have non-overlapping groups, recursively partitioning the data until each partition is either pure or sufficiently meet a requirement, e.g., a parameter set by the user, and using function `value (A) < x` to analyze attributes (Shafer, page 545:2 to 546:1). As seen, `value (A) < x` is ***a hard cluster analysis*** for building the decision tree.

(4) As argued by appellants at page 13 with respect to claims 13 and 30:

> *Claim 13 dependent on claim 10. Since Janikow's separation of classification and tree-building teaches against claim 10, Choe et al.'s different classification function does not undo Janikow's teaching against.*

Examiner respectfully traverses because Janikow does not teach against claim 10 as discussed above. Therefore, Janikow method could apply fuzzy c-means analyses as taught by Choe to maximize the number of data points in a cluster.

As argued by appellants at page 13 with respect to claims 14 and 31:

> *Janikow's separation of classification and tree-building teaches against claim 10, upon which claim 14 depends, and Shafer et al.'s different classification function does not undo Janikow's teaching against.*

Examiner respectfully traverses because Janikow does not teach against claim 10 as discussed above. Therefore, Janikow method could apply hard cluster analysis as taught by Shafer for classifying records of unknown class.

C. RESPONSE TO ARGUMENTS WITH RESPECT TO THE REJECTION OF CLAIMS 1-3 AND 18-22 UNDER 35 U.S.C § 103 AS BEING UNPATENABLE OVER RASTOGI AND SHIMOJI.

(1) As argued by appellants at page 13 and 14 with respect to claims 1-3 and 18-22:

> *Rastogi et al. in view of Shimoji et al. fail to disclose the limitations of these claims. For example, independent claims 1 and 18 recite: "performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters."*
> *Nowhere does Rastogi et al. describe "cluster analysis" or even a split based on any type of cluster analysis. In fact, Rastogi et al. nowhere mentions a "cluster."*

Examiner respectfully traverses. As shown in FIG. 1 of Rastogi, a single record corresponding to each loan request characterizes two attributes, salary and education level completed (Col. 2, lines 50-56). As shown in FIG. 2, `salary` is selected from among the features characterizing the data associated with the root node, and the test `is the salary level of the applicant less than $20,000.00` (Col. 2, lines 62-63) is to split the root node N into N_1 and N_2 (FIG. 3, line 8). The test is based on the process of calculation of the least entropy by scanning the attribute list from the beginning to calculate an entropy for each split point or each numeric attribute in order to determine the least entropy (Col. 4, lines 25-52). As seen, the process of determining the least entropy as ***performing a cluster analysis along the selected feature to group the data into one or more cluster***, e.g., data of root node N is grouped into N_1 and N_2. The test `is the`

salary level of the applicant less than $20,000.00 based on the calculated entropy describes ***a cluster analysis*** to split the root node into N_1 and N_2, which are *clusters*. Rastogy does not teach the claimed ***distances between the data and respective one of more centers of the one or more clusters*** is used for cluster analysis. Shimoji discloses a method of clustering a set of data by using a clustering error ***based on distances between the data and respective one of more centers of the one or more clusters*** (Shimoji, Introduction). Thus, instead of entropy, ***distances between the data and respective one of more centers of the one or more clusters*** can be used for the test *is the salary level of the applicant less than $20,000.00.*

(2) As argued by appellants at page 15 with respect to claims 1-3 and 18-22:

> *Nowhere does Shimoji et al. disclose or suggest "performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters."*

Examiner respectfully traverses because the process of performing a cluster analysis along the selected feature is disclosed by Rastogi as discussed above. The missing of Rastogi is a distance for supporting the clustering process. There is no need of disclosing a cluster analysis from Shimoji as argued by appellants.

As argued by appellants at page 15 with respect to claims 1-3 and 18-22:

> *Thus, there is no motivation to combine Rastogi et al. and Shimoji et al., other than impermissible hindsight. Thus, the rejection of claims 1-3 and 18-22 based on Rastogi et al. in view of Shimoji et al. should be withdrawn.*

In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988), and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, both Rastogi and Shimoji disclose the technique of training data, and the detail of clustering technique as taught by Shimoji is a must for Rastogi.

In response to applicant's argument that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

(3) As argued by appellants at page 16 with respect to claims 4 and 21:

The rejection of claims 4 and 21 based on Rastogi et al., Shimoji et al., and Background should also be reversed. The Background does not cure the factual deficiencies or the lack of motivation to combine.

In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988), and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, both Rastogi and Shimoji disclose the technique of training data, and the detail of clustering technique as taught by Shimoji is a must for Rastogi.

(4) As argued by appellants at page 16 with respect to claims 5 and 22:

> *As a result, there is no teaching or suggesting in Hall et al., of recursively performing the cluster analysis while "refining a node of a decision tree."*

In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., *recursively performing the cluster analysis while "refining a node of a decision tree"*) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

(5) Appellants' argument at page 17 about Shafer reference is respectfully traverses with the reasons as discussed in claims 6 and 23 above.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

HUNG Q PHAM
Examiner
Art Unit 2162

June 8, 2005

Conferees

SAFET METJAHIC, Supervisory Patent Examiner, AU 2161.
HOSAIN ALAM, Supervisory Paten Examiner, AU 2166.

Stephen C Carlson
Ditthavong & Carlson P C
10507 Braddock Road
Suite A
Fairfax, VA 22032

HOSAIN ALAM
PATENT EXAMINER

SAFET METJAHIC
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100